ИНФОРМАТИКА, ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА И УПРАВЛЕНИЕ
INFORMATION TECHNOLOGY, COMPUTER SCIENCE, AND MANAGEMENT



УДК 519.7
10.23947/1992-5980-2018-18-1-92-101

О применимости математического масштабирования и нормирования при решении прикладных задач[*]

А. И. Долгов[1], Д. В. Маршаков[2]**
[1] АО «Всероссийский научно-исследовательский институт «Градиент», г. Ростов-на-Дону, Российская Федерация
[2] Донской государственный технический университет, г. Ростов-на-Дону, Российская Федерация

On applicability of mathematical scaling and normalization in applied problem solving [***]

A. I. Dolgov[1], D. V. Marshakov[2]**
[1] All-Russian Scientific Research Institute “Gradient” JSC, Rostov-on-Don, Russian Federation
[2] Don State Technical University, Rostov-on-Don, Russian Federation

Введение. Статья посвящена анализу применимости математического масштабирования и нормирования при решении различных прикладных задач. Рассмотрены наиболее известные формулы, часто используемые в ходе теоретических и практических изысканий. Цель работы — выявление свойств математического масштабирования и нормирования.
Материалы и методы. На конкретных вычислительных примерах оценены ошибки, получаемые при использовании формул математического масштабирования и нормирования. С учетом сравнительной оценки отношений степени величины исходных и результирующих значений (а также отношений степени их различия) оценивается корректность получаемых данных, оказывающих существенное влияние на результирующие значения.
Результаты исследования. Проведенный анализ позволяет сделать выводы о том, что известные формулы математического масштабирования и нормирования обладают свойствами, не учитывавшимися в теории и практике.
Обсуждение и заключения. Полученные результаты позволяют избежать ошибочных решений, обусловленных использованием неприменимых формул масштабирования и нормирования при решении задач в теории и практике экономики, организационного управления, медицины и многих других областей.

Introduction. The applicability of mathematical scaling and normalization in solving various applied problems is analyzed. The best known formulas often used along the theoretical and experimental studies are considered. The purpose of this work is to identify the properties of mathematical scaling and rationing.
Materials and Methods. The errors obtained under using the mathematical scaling and normalization formulas are considered via specific computational examples. Based on a comparative evaluation of the ratio of the degree of magnitude of the initial and resulting values (as well as the ratio of the degree of difference of these values), the correctness of the results obtained which significantly effects the final values is estimated.
Research Results. The analysis leads to the conclusion that some known mathematical scaling and normalization formulas possess properties that are ignored in theory and practice.
Discussion and Conclusions. The results obtained allow avoiding erroneous decisions caused by the use of invalid scaling and normalization formulas under solving problems in theory and practice of economics, administrative management, medicine, and plenty of other fields.

Ключевые слова: масштабирование, нормирование, анализ данных, применимость формул, искусственная нейронная сеть, формула Байеса.

Keywords: scaling, normalization, data analysis, applicability of formulas, artificial neural network, Bayes’ rule.

Образец для цитирования: Долгов, А. И. О применимости математического масштабирования и нормирования / А. И. Долгов, Д. В. Маршаков // Вестник Дон. гос. техн. ун-та. — 2018. — Т. 18, № 1. — С. 92−101.

For citation: A.I. Dolgov, D.V. Marshakov. On applicability of mathematical scaling and normalization in applied problem solving. Vestnik of DSTU, 2018, vol. 18, no.1, pp. 92−101.

[*] Работа выполнена в рамках инициативной НИР.
[**] E-mail: dolgov-ai@yandex.ru, daniil_marshakov@mail.ru
[***] The research is done within the frame of the independent R&D.

Введение. Важным этапом обработки информации является приведение исходных данных к допустимым в системе интервалам числовых значений [1–4]. При этом принято говорить о двух общих подходах, реализующих данное преобразование, — масштабировании и нормировании. Эти термины часто довольно свободно используются в различных областях, и смысл их определяется контекстом.

В имеющейся литературе речь идет о разнообразных объектах масштабирования. В их числе: топографические карты, планы, рисунки, чертежи для проектирования, интернет-страницы. Кроме того, масштабируются бизнес- и производственные процессы — например, нормирование труда, капитала, времени, материалов и других ресурсов. Безусловно, при масштабировании и нормировании широко применяются математические методы и средства вычислительной техники.

В данной публикации рассматриваются виды математического масштабирования и нормирования, использующие широко применяемые формулы. Следует отметить, что не относятся к рассматриваемым видам масштабирование и нормирование матриц, логарифмическое масштабирование и нормирование, а также масштабирование и нормирование с применением коэффициентов, являющихся множителями.

С учетом сказанного, рассматриваемые виды математического масштабирования и нормирования предполагают, что:
— изменяются исходные данные, представленные в виде числовых значений или аналитических выражений (и определяемые ими);
— получаются результирующие данные, принадлежащие новому интервалу числовых величин.

Обосновывается утверждение о том, что изменение исходных данных, приводящее к другому интервалу их значений, в некоторых случаях зависит от постановки задачи. При этом формулировка проблемы может быть неоднозначной и приводить к двояко трактуемым результатам, понимаемым как итог математического масштабирования либо нормирования. Например, математическое нормирование может рассматриваться как частный случай математического масштабирования при приведении исходных данных в числовой промежуток [0,1] [4].

В данной работе решается задача анализа применимости математического масштабирования и нормирования в некоторых известных формулах [5–9]. Оценивается корректность получаемых данных, существенно влияющих на результирующие значения, используемые в теории и практике экономики, управления, медицины и многих других областях. При этом учитываются:
— сравнительная оценка отношений степени величины исходных и результирующих значений, вычисляемых с применением известных формул;
— отношения степени различия таких значений.

Материалы и методы. Основные определения. Для правильной трактовки сущности и применимости формул математического масштабирования и нормирования используем известные определения.

Масштабировать:
1) изменять масштаб, размах чего-либо;
2) делать более масштабным что-либо;
2) изменять масштаб, представлять в другом масштабе (об изображениях, шрифтах и т. п.).

Нормировать:
1) установить пределы чего-либо;
2) ввести норму в чем-либо (вариант: ввести в норму что-либо);
2) устанавливать законные пределы чего-нибудь [10, 11].

Проводимый далее анализ ограничивается рассмотрением формул:
— реализующих математическое масштабирование или нормирование в соответствии с подходами, представленными в [3–9];
— использованных при иллюстрации примеров решения практических задач.

Такие формулы, обеспечивающие изменение значений исходных числовых данных $X=\bigcup_i^m x_i=\{x_1,x_2,\ldots,x_n\}$ и приведение их к новому интервалу числовых значений $\tilde{x}_i$, записываются в следующем виде:

$$\tilde{x}_i=\frac{x_i}{\max X}, \tag{1}$$

$$\tilde{x}_i=\frac{x_i}{\sum_{i=1}^{n} x_i}, \tag{2}$$

$$\tilde{x}_i = \frac{x_i - \min X}{\max X - \min X}\,, \tag{3}$$

$$\tilde{x}_i = \frac{2(x_i - \max X)}{\max X - \min X} = 2\frac{x_i - \min X}{\max X - \min X} - 1\,, \tag{4}$$

где $\max X$ и $\min X$ — наибольшее и наименьшее из числовых значений, указанных в заданных исходных данных.

С учетом приведенных ранее определений понятий масштабирования и нормирования формулы (1)–(4) могут быть получены при двух разных постановках задачи.

1) Математическое масштабирование: выполнить изменение исходных числовых значений x_i с применением общего делителя D для получения принадлежащих другому интервалу результирующих числовых значений $\tilde{x}_i$.

В результате решения этой задачи могут быть получены:
— при $D = \max X$ формула (1), аналогичная указанной в [5];
— при $D = \sum_{i=1}^{n} x_i$ формула (2), аналогичная использованным в [6, 7];
— при $D = (\max X - \min X)$ формулы (3) и (4), в зависимости от заданных интервалов использованные в [3, 8, 9] соответственно, а также аналогичные указанным в [5].

2) Математическое нормирование: выполнить изменение исходных числовых значений x_i с применением общего делителя и получением результирующих числовых значений $\tilde{x}_i$, принадлежащих задаваемому новому интервалу.

В результате решения такой задачи могут быть получены:
— при интервале $\left[\frac{\max X}{\max X}, \frac{\min X}{\max X}\right] = \left[1, \frac{\min X}{\max X}\right]$ — формула (1);
— при интервале $\left[\max X \Big/ \sum_{i=1}^{n} x_i, \min X \Big/ \sum_{i=1}^{n} x_i\right]$ — формула (2);
— при интервале [0,1] — формула (3);
— при интервале [–1, 1] — формула (4).

Анализ интернет-источников демонстрирует отсутствие унифицированного подхода к рассматриваемой терминологии. Изменения исходных данных с использованием математических формул (1)–(4) могут называться как масштабированием, так и нормированием. Общие определения этих понятий отсутствуют, при этом можно найти частные формулировки.

Следует отметить весьма важное обстоятельство. Применимость формулы математического нормирования и масштабирования зависит от равенства отношений:
— величины исходных и результирующих значений,
— степеней различия исходных и результирующих значений.

Задача данной работы формулируется следующим образом. Исследуются заданные исходные данные. На основе этого анализа оценивается применимость формулы математического масштабирования или нормирования числителя на предмет отсутствия степеней различия отношений исходного x_i и результирующего $\tilde{x}_i$ значения.

Результаты исследования. Анализ применимости. Применяя формулы (1)–(4) для трех исходных числовых значений $x_j, x_k, x_l \in \bigcup_{i}^{n} x_i$, выполним аналитическое сопоставление данных в соответствии с задачей исследования. Итак, сравним отношения степеней величины:
— результирующих значений $\frac{\tilde{x}_j}{\tilde{x}_k}$, $\frac{\tilde{x}_k}{\tilde{x}_l}$ и степеней их различия $\frac{\tilde{x}_j - \tilde{x}_k}{\tilde{x}_k - \tilde{x}_l}$;
— исходных значений $\frac{x_j}{x_k}$, $\frac{x_k}{x_l}$ и степеней их различия $\frac{x_j - x_k}{x_k - x_l}$.

1) Формула (1): $\tilde{x}_j = \frac{x_j}{\max X}$, $\tilde{x}_k = \frac{x_k}{\max X}$, $\tilde{x}_l = \frac{x_l}{\max X}$;

отношения степеней величины результирующих значений:

$$\frac{\tilde{x}_j}{\tilde{x}_k} = \frac{x_j}{\max X} \Big/ \frac{x_k}{\max X} = \frac{x_j}{x_k}, \qquad \frac{\tilde{x}_k}{\tilde{x}_l} = \frac{x_k}{\max X} \Big/ \frac{x_l}{\max X} = \frac{x_k}{x_l};$$

отношения степеней различия результирующих значений:

$$\frac{\tilde{x}_j - \tilde{x}_k}{\tilde{x}_k - \tilde{x}_l} = \left(\frac{x_j}{\max X} - \frac{x_k}{\max X} \right) \Big/ \left(\frac{x_k}{\max X} - \frac{x_l}{\max X} \right) = \frac{x_j - x_k}{x_k - x_l}.$$

2) Формула (2): $\tilde{x}_j = \frac{x_j}{\sum_{i=1}^{n} x_i}$, $\tilde{x}_k = \frac{x_k}{\sum_{i=1}^{n} x_i}$, $\tilde{x}_l = \frac{x_l}{\sum_{i=1}^{n} x_i}$;

отношения степеней величины результирующих значений:

$$\frac{\tilde{x}_j}{\tilde{x}_k} = \frac{x_j}{\sum_{i=1}^{n} x_i} \Big/ \frac{x_k}{\sum_{i=1}^{n} x_i} = \frac{x_j}{x_k}, \qquad \frac{\tilde{x}_k}{\tilde{x}_l} = \frac{x_k}{\sum_{i=1}^{n} x_i} \Big/ \frac{x_l}{\sum_{i=1}^{n} x_i} = \frac{x_k}{x_l};$$

отношения степеней различия результирующих значений:

$$\frac{\tilde{x}_j - \tilde{x}_k}{\tilde{x}_k - \tilde{x}_l} = \left(\frac{x_j}{\sum_{i=1}^{n} x_i} - \frac{x_k}{\sum_{i=1}^{n} x_i}\right) \Big/ \left(\frac{x_k}{\sum_{i=1}^{n} x_i} - \frac{x_l}{\sum_{i=1}^{n} x_i}\right) = \frac{x_j - x_k}{x_k - x_l}.$$

3) Формула (3):

$$\tilde{x}_j = \frac{x_j - \min X}{\max X - \min X}, \ \tilde{x}_k = \frac{x_k - \min X}{\max X - \min X}, \ \tilde{x}_l = \frac{x_l - \min X}{\max X - \min X};$$

отношения степеней величины результирующих значений:

$$\frac{\tilde{x}_j}{\tilde{x}_k} = \left(\frac{x_j - \min X}{\max X - \min X} \right) \Big/ \left(\frac{x_k - \min X}{\max X - \min X} \right) = \frac{x_j - \min X}{x_k - \min X} \neq \frac{x_j}{x_k};$$

$$\frac{\tilde{x}_k}{\tilde{x}_l} = \left(\frac{x_k - \min X}{\max X - \min X} \right) \Big/ \left(\frac{x_l - \min X}{\max X - \min X} \right) = \frac{x_k - \min X}{x_l - \min X} \neq \frac{x_k}{x_l};$$

отношения степеней различия результирующих значений:

$$\frac{\tilde{x}_j - \tilde{x}_k}{\tilde{x}_k - \tilde{x}_l} = \frac{\dfrac{x_j - \min X}{\max X - \min X} - \dfrac{x_k - \min X}{\max X - \min X}}{\dfrac{x_k - \min X}{\max X - \min X} - \dfrac{x_l - \min X}{\max X - \min X}} = \frac{x_j - x_k}{x_k - x_l}.$$

4) Формула (4):

$$\tilde{x}_j = 2\frac{x_j - \min X}{\max X - \min X} - 1, \ \tilde{x}_k = 2\frac{x_k - \min X}{\max X - \min X} - 1, \ \tilde{x}_l = 2\frac{x_l - \min X}{\max X - \min X} - 1;$$

отношения степеней величины результирующих значений:

$$\frac{\tilde{x}_j}{\tilde{x}_k} = \left(2\frac{x_j - \min X}{\max X - \min X} - 1 \right) \Big/ \left(2\frac{x_k - \min X}{\max X - \min X} - 1 \right) =$$

$$= \left(\frac{2x_j - 2\min X - \max X + \min X}{\max X - \min X} \right) \Big/ \left(\frac{2x_k - 2\min X - \max X + \min X}{\max X - \min X} \right) = \frac{2x_j - \min X - \max X}{2x_k - \min X - \max X} \neq \frac{x_j}{x_k};$$

аналогично $\frac{\tilde{x}_k}{\tilde{x}_l} = \frac{2x_k - \min X - \max X}{2x_l - \min X - \max X} \neq \frac{x_k}{x_l}$;

отношения степеней различия результирующих значений:

$$\frac{\tilde{x}_j - \tilde{x}_k}{\tilde{x}_k - \tilde{x}_l} = \frac{\left(2\dfrac{x_j - \min X}{\max X - \min X} - 1 \right) - \left(2\dfrac{x_k - \min X}{\max X - \min X} - 1 \right)}{\left(2\dfrac{x_k - \min X}{\max X - \min X} - 1 \right) - \left(2\dfrac{x_l - \min X}{\max X - \min X} - 1 \right)} =$$

$$=\frac{\left(2x_j-\min X-\max X\right)-\left(2x_k-\min X-\max X\right)}{\left(2x_k-\min X-\max X\right)-\left(2x_l-\min X-\max X\right)}=\frac{2x_j-2x_k}{2x_k-2x_l}=\frac{x_j-x_k}{x_k-x_l}.$$

На примере выражения (2) рассмотрим неприменимость формул, реализующих математическое масштабирование в случае использования делителя, не являющегося общим для сравниваемых и (или) совместно обрабатываемых числовых значений.

При постановке в формулу (2) значений $x_i=P\left(H_i\right)P\left(E\,|\,H_i\right)$ и $\tilde{x}_i=P\left(E\,|\,H_i\right)$, где $\bigcup_i^m x_i=\left\{x_1,x_2,\ldots,x_j,\ldots,x_n\right\}$, получается общеизвестная [6, 7, 12, 13] формула Байеса:

$$P\left(H_i\,|\,E\right)=\frac{P\left(H_i\right)P\left(E\,|\,H_i\right)}{\sum_{i=1}^{n}P\left(H_i\right)P\left(E\,|\,H_i\right)}. \quad (5)$$

Представляется интересным анализ модификации формулы Байеса, использованной при решении задач построения экспертной системы в области медицинской диагностики [14] и создания диагностирующей экспертной системы для выбора компьютера пользователем [1]:

$$P\left(H_j\,|\,E\right)=\frac{P\left(H_j\right)P\left(E\,|\,H_j\right)}{P\left(H_j\right)P\left(E\,|\,H_j\right)+P\left(неH_j\right)P\left(E\,|\,неH_j\right)}, \quad (6)$$

где $P\left(E\,|\,неH_j\right)$ — заблаговременно определяемая наряду с априорной условной вероятностью $P\left(E\,|\,H_j\right)$ априорная условная вероятность наличия события E при отрицании гипотезы H_j.

Формула (6), названная в публикациях [1] и [14] формулой Байеса, отличается от нее выражением в знаменателе и предложена в публикации [14] в качестве весьма удобной для рекуррентных вычислений. Как отмечает автор [14] (далее в кавычках приводятся цитируемые фрагменты в переводе с английского), в случае накапливания свидетельств, вычислив апостериорную вероятность гипотезы H_k для одного учитываемого свидетельства E, «мы забываем об этом, за исключением того, что априорная вероятность $P\left(H\right)$ заменяется на $P\left(H\,|\,E\right)$». Затем продолжается выполнение вычислений, «но с учетом постоянной коррекции значения $P\left(H\right)$ по мере поступления новой информации» (свидетельств).

Необходимо отметить, что формула (6) может быть названа формулой Байеса лишь в том случае, если ее знаменатель равен знаменателю, указанному в формуле (5), то есть при

$$P\left(H_j\right)P\left(E\,|\,H_j\right)+P\left(неH_j\right)P\left(E\,|\,неH_j\right)=\sum_{i=1}^{n}P\left(H_i\right)P\left(E\,|\,H_i\right).$$

Следует иметь в виду, что, в отличие от (5), в (6), используется дополнительное априорное значение $P\left(E\,|\,неH_j\right)$, которое при условии равенства знаменателей может быть определено из уравнения:

$$P\left(H_j\right)P\left(E\,|\,H_j\right)+P\left(неH_j\right)P\left(E\,|\,неH_j\right)=\sum_{i=1}^{n}P\left(H_i\right)P\left(E\,|\,H_i\right).$$

При этом

$$P\left(E\,|\,неH_j\right)=\frac{\sum_{i=1(i\neq j)}^{n}P\left(H_i\right)P\left(E\,|\,H_i\right)}{P\left(неH_j\right)}=\frac{\sum_{i=1(i\neq j)}^{n}P\left(H_i\right)P\left(E\,|\,H_i\right)}{1-P\left(H_j\right)}.$$

Применив формулу (2) для трех значений исходных данных $x_j=P\left(H_j\right)P\left(E\,|\,H_j\right)$, $x_k=P\left(H_k\right)P\left(E\,|\,H_k\right)$, $x_l=P\left(H_l\right)P\left(E\,|\,H_l\right)$, выполним аналитическое сравнение отношений степеней величины:

— результирующих значений $\frac{P\left(E\,|\,неH_j\right)}{P\left(E\,|\,неH_k\right)}$, $\frac{P\left(E\,|\,неH_k\right)}{P\left(E\,|\,неH_l\right)}$ и степеней их различия $\frac{P\left(E\,|\,неH_j\right)-P\left(E\,|\,неH_k\right)}{P\left(E\,|\,неH_k\right)-P\left(E\,|\,неH_l\right)}$;

— исходных значений $\frac{P\left(H_j\right)P\left(E\,|\,H_j\right)}{P\left(H_k\right)P\left(E\,|\,H_k\right)}$, $\frac{P\left(H_k\right)P\left(E\,|\,H_k\right)}{P\left(H_l\right)P\left(E\,|\,H_l\right)}$ и степеней их различия $\frac{P\left(H_j\right)P\left(E\,|\,H_j\right)-P\left(H_k\right)P\left(E\,|\,H_k\right)}{P\left(H_k\right)P\left(E\,|\,H_k\right)-P\left(H_l\right)P\left(E\,|\,H_l\right)}$.

Ограничимся сравнением апостериорных условных вероятностей, получаемых при одной и той же гипотезе, использованной в публикациях [1] и [14], с применением формул (5) и (6). Так как в (5) в отличие от (6)

не учитывается дополнительно определяемая априорная условная вероятность $P\left(E \mid неH_j\right)$, то правые части формул в общем случае не могут быть равными:

$$P\left(H_j \mid E\right)=\frac{P\left(H_j\right)P\left(E \mid H_j\right)}{\sum_{i=1}^{n}P\left(H_i\right)P\left(E \mid H_i\right)} \neq \frac{P\left(H_j\right)P\left(E \mid H_j\right)}{P\left(H_j\right)P\left(E \mid H_j\right)+P\left(неH_j\right)P\left(E \mid неH_j\right)} .$$

Если полагать, что широко применяемая формула Байеса приводит к правильным результатам, то формула (6) должна быть признана неприменимой.

Подводя итог сказанному, следует констатировать наличие у рассмотренных формул математического масштабирования и нормирования свойств, существенно влияющих на их применимость. При этом ранее не учитывались перечисленные ниже факты.

1) Формулы (1) и (2) применимы при решениях задач сравнительной оценки или совместной обработки числовых данных, и правильные результаты получаются при оценке степени как результирующих величин, так и их различия.

2) Формулы (3) и (4) применимы только при решениях задач, учитывающих степени различия результирующих значений. Использование их для решения задач, учитывающих степени величин результатов, приводит к существенным ошибкам.

3) В некоторых случаях применение рассмотренных формул недопустимо, т. к. приводит к некорректному результату. Речь идет о математическом нормировании с использованием делителя, не являющегося общим для сравниваемых и (или) совместно обрабатываемых исходных данных. Таковым является, например, нормирование модификаций формулы Байеса (5) с совместно обрабатываемыми априорными исходными данными в числителях и знаменателях, в частности записанных в виде (6).

Сравнительные оценки применимых и неприменимых формул. Теоретические выводы о случаях применимости и неприменимости формул становятся более убедительными, если подтверждаются наглядными примерами вычислений, используемых при решении конкретных прикладных задач.

1. Рассмотрим классическую задачу классификации «Ирисы Фишера» (см., например, [2, 8, 15]) с применением математического масштабирования исходных данных.

Даны исходные показатели измерений трех видов цветков ириса: *setosa*, *versicolor* и *virginica*. Всего имеется по 50 экземпляров каждого вида. Для каждого из 150 экземпляров измерены четыре величины: длина и ширина чашелистика, длина и ширина лепестка. Требуется определить тип нового цветка по исходным данным, полученным при проведении измерений.

Задача интересна тем, что одним из способов ее решения является применение искусственной нейронной сети с выбором вида масштабирования исходных данных.

Рассматриваемые примеры масштабируемых исходных данных, обычно используемых при обучении искусственной нейронной сети, частично представлены в табл. 1.

Таблица 1
Table 1

Фрагмент набора данных «Ирисы Фишера»

Fisher's Iris data set fragment

№	Длина чашелистика	Ширина чашелистика	Длина лепестка	Ширина лепестка	1-й сорт *setosa*	2-й сорт *versicolor*	3-й сорт *virginica*
1	5,1	3,5	1,4	0,2	1	0	0
…	…	…	…	…	…	…	…
38	4,9	3,6	1,4	0,1	1	0	0
…	…	…	…	…	…	…	…
132	7,9	3,8	6,4	2	0	0	1
…	…	…	…	…	…	…	…
150	5,9	3	5,1	1,8	0	0	1

В частично представленном наборе данных «Ирисы Фишера» минимальное и максимальное значения величин полного набора $\min X = 0{,}1$ и $\max X = 7{,}9$ указаны соответственно в строках 38 и 132.

Выражения (1) и (2) вполне применимы для оценки как степеней величины, так и степеней различия масштабированных исходных данных, поэтому для анализа масштабирования используем формулы (3) и (4).

1) Определяемые по (3) результаты математического масштабирования исходных данных, указанных в 1-й и 150-й строках таблицы:

$$\frac{5{,}1-0{,}1}{7{,}9-0{,}1}=\frac{5}{7{,}8}=0{,}6410\,;\ \frac{3{,}5-0{,}1}{7{,}9-0{,}1}=\frac{3{,}4}{7{,}8}=0{,}4359\,;$$

$$\frac{1{,}4-0{,}1}{7{,}9-0{,}1}=\frac{1{,}3}{7{,}8}=0{,}1705\,;\ \frac{0{,}2-0{,}1}{7{,}9-0{,}1}=\frac{0{,}1}{7{,}8}=0{,}0128\,;$$

$$\frac{5{,}9-0{,}1}{7{,}9-0{,}1}=\frac{5{,}8}{7{,}8}=0{,}7436\,;\ \frac{3-0{,}1}{7{,}9-0{,}1}=\frac{2{,}9}{7{,}8}=0{,}3718\,;$$

$$\frac{5{,}1-0{,}1}{7{,}9-0{,}1}=\frac{5}{7{,}8}=0{,}6410\,;\ \frac{1{,}8-0{,}1}{7{,}9-0{,}1}=\frac{1{,}7}{7{,}8}=0{,}2179\,.$$

При этом отношения исходных данных, указанных в 1-й и 150-й строках частично представленного набора данных (записываемые с округлением до четырех знаков дробной части), равны:

$$\frac{5{,}1}{5{,}9}=0{,}8644\,;\quad \frac{3{,}5}{3}=1{,}1666(6)\,;\quad \frac{1{,}4}{5{,}1}=0{,}2745\,;\quad \frac{0{,}2}{1{,}8}=0{,}1111(1)\,.$$

Отношения соответствующих масштабированных данных:

$$\frac{5}{7{,}8}\Big/\frac{5{,}8}{7{,}8}=\frac{5}{5{,}8}=0{,}8621\,;\qquad \frac{3{,}4}{7{,}8}\Big/\frac{2{,}9}{7{,}8}=\frac{3{,}4}{2{,}9}=1{,}1724\,;$$

$$\frac{1{,}3}{7{,}8}\Big/\frac{5}{7{,}8}=\frac{1{,}3}{5}=0{,}26\,;\qquad \frac{0{,}1}{7{,}8}\Big/\frac{1{,}7}{7{,}8}=\frac{0{,}1}{1{,}7}=0{,}0588\,.$$

Разница есть и бывает существенной — например, в случае длины лепестка отличие $\frac{1{,}4}{5{,}1}\Big/\frac{\frac{1{,}3}{7{,}8}}{\frac{5}{7{,}8}}=1{,}6471$, а в случае ширины лепестка $\frac{0{,}2}{1{,}8}\Big/\frac{\frac{0{,}1}{7{,}8}}{\frac{1{,}7}{7{,}8}}=1{,}888(8)$, то есть более чем в полтора раза.

2) Определяемые по формуле (4) результаты масштабирования исходных данных, указанных в 1-й и 150-й строках таблицы:

$$2\frac{5{,}1-0{,}1}{7{,}9-0{,}1}-1=\frac{10}{7{,}8}-1=0{,}2821;\quad 2\frac{3{,}5-0{,}1}{7{,}9-0{,}1}-1=\frac{6{,}8}{7{,}8}-1=-0{,}1282\,;$$

$$2\frac{1{,}4-0{,}1}{7{,}9-0{,}1}-1=\frac{2{,}6}{7{,}8}-1=-0{,}6666(6)\,;\quad 2\frac{0{,}2-0{,}1}{7{,}9-0{,}1}-1=\frac{0{,}2}{7{,}8}-1=-0{,}9487\,;$$

$$2\frac{5{,}9-0{,}1}{7{,}9-0{,}1}-1=\frac{11{,}6}{7{,}8}-1=0{,}4872\,;\quad 2\frac{3-0{,}1}{7{,}9-0{,}1}-1=\frac{5{,}8}{7{,}8}-1=-0{,}2564\,;$$

$$2\frac{5{,}1-0{,}1}{7{,}9-0{,}1}-1=\frac{5}{7{,}8}-1=0{,}2821;\quad 2\frac{1{,}8-0{,}1}{7{,}9-0{,}1}-1=\frac{1{,}7}{7{,}8}-1=-0{,}5641\,.$$

При этом отношения исходных данных, указанных в 1-й и 150-й строках фрагмента таблицы, равны

$$\frac{5{,}1}{5{,}9}=0{,}8644\,;\quad \frac{3{,}5}{3}=1{,}1666(6)\,;\quad \frac{1{,}4}{5{,}1}=0{,}2745\,;\quad \frac{0{,}2}{1{,}8}=0{,}1111(1)\,.$$

Отношения соответствующих масштабированных данных:

$$\left(\frac{10}{7{,}8}-1\right)\Big/\left(\frac{11{,}6}{7{,}8}-1\right)=\frac{2{,}2}{3{,}8}=0{,}5789\,;\qquad \left(\frac{3{,}4}{7{,}8}-1\right)\Big/\left(\frac{2{,}9}{7{,}8}-1\right)=\frac{-4{,}4}{-5{,}1}=0{,}8627\,;$$

$$\left(\frac{2{,}6}{7{,}8}-1\right)\Big/\left(\frac{5}{7{,}8}-1\right)=\frac{5{,}2}{2{,}8}=1{,}8571\,;\qquad \left(\frac{0{,}2}{7{,}8}-1\right)\Big/\left(\frac{1{,}7}{7{,}8}-1\right)=\frac{-7{,}6}{-6{,}1}=1{,}2459\,.$$

В случае формулы (4) разница оказывается существенно больше. Например, отличие отношений масштабированных и исходных значений длины лепестка $\left[\left(\frac{2,6}{7,8}-1\right)\Big/\left(\frac{5}{7,8}-1\right)\right]\Big/\frac{1,4}{5,1}=6,7653$, а отличие отношений исходных и масштабированных значений ширины лепестка $\frac{0,2}{1,8}\Big/\left[\left(\frac{0,2}{7,8}-1\right)\Big/\left(\frac{1,7}{7,8}-1\right)\right]=11,2131$, то есть более чем в 10 раз.

2. Решим задачу сравнительной оценки результатов выбора компьютера покупателем-пользователем. Применим математическое нормирование исходных данных [1], а также формулы (6) и (5). Рассмотрим упрощенный пример определения апостериорной условной вероятности двух гипотез при учете одного свидетельства.

В процессе решения задачи определяются апостериорные условные вероятности двух гипотез: H_1 — выбор компьютера *Game Professional* либо H_2 — выбор компьютера *Game Master*. Учитывается одно свидетельство E в виде ответа на вопрос о потребностях пользователя.

Исходные данные: $P(H_1)=0,08$; $P(H_2)=0,15$; $P(E\,|\,H_1)=0,4$; $P(E\,|\,H_2)=0,3$; $P(E\,|\,неH_1)=0,03$; $P(E\,|\,неH_2)=0,1$.

Результаты, определяемые по (6):

$$P(H_1\,|\,E)=\frac{P(H_1)P(E\,|\,H_1)}{P(H_1)P(E\,|\,H_1)+P(неH_1)P(E\,|\,неH_1)}=\frac{0,08\times0,4}{0,08\times0,4+0,92\times0,03}=0,5369;$$

$$P(H_2\,|\,E)=\frac{P(H_2)P(E\,|\,H_2)}{P(H_2)P(E\,|\,H_2)+P(неH_2)P(E\,|\,неH_2)}=\frac{0,15\times0,3}{0,15\times0,3+0,85\times0,1}=0,3462.$$

Результаты, определяемые по (5):

$$P(H_1\,|\,E)=\frac{P(H_1)P(E\,|\,H_1)}{P(H_1)P(E\,|\,H_1)+P(H_2)P(E\,|\,H_2)}=\frac{0,08\times0,4}{0,08\times0,4+0,15\times0,3}=0,1063;$$

$$P(H_2\,|\,E)=\frac{P(H_2)P(E\,|\,H_2)}{P(H_1)P(E\,|\,H_1)+P(H_2)P(E\,|\,H_2)}=\frac{0,15\times0,3}{0,08\times0,4+0,15\times0,3}=0,1495.$$

Сравнительная оценка результатов, полученных с применением формул (6) и (5), позволяет сделать удивительные выводы.

1) Использование формулы (5) с корректным масштабированием приводит к решению о целесообразности выбора компьютера *Game Master*, ввиду того, что

$$P(H_2\,|\,E)=0,1495>P(H_1\,|\,E)=0,1063.$$

2) Использование формулы (6) с неодинаковыми знаменателями приводит к противоположному, ошибочному заключению — выбору компьютера *Game Professional*, так как

$$P(H_1\,|\,E)=0,5369>P(H_2\,|\,E)=0,3462.$$

Из этого следуют два заключения.

1) Нормирование, использованное в формуле (5), названной в [1] и [14] формулой Байеса, неприменимо, так как имеет два недостатка:

— априорная условная вероятность $P(E\,|\,неH_k)$ является задаваемой, в отличие от существующего варианта формулы Байеса с общим делителем с вычисляемой апостериорной вероятностью $P(E\,|\,неH_k)$ на основе априорных исходных данных $P(H_k)$, $P(E\,|\,H_k)$;

— при разных числителях отличаются и знаменатели, то есть отсутствует общий делитель.

2) При сравнительной оценке и (или) совместной обработке масштабируемых значений, получаемых с использованием (5) с задаваемой дополнительной априорной условной вероятностью $P(E\,|\,неH_k)$, из-за неодинаковых знаменателей возникают ошибки, возрастающие в процессе последовательного учета каждого очередного свидетельства.

Обсуждение и заключения. Проведенный анализ позволяет утверждать, что рассмотренные формулы обладают свойствами, не учитывавшимися в теории и практике математического масштабирования и нормирования. В частности, можно сделать ряд выводов.

1) Формулы, в числителях которых указывается только одно из исходных значений исходных данных, обеспечивают получение корректных результатов и применимы во всех случаях математического масштабирования и нормирования.

2) Формулы, в числителях которых вычитается минимальное значение исходных данных, являются применимыми в случае сравнительной оценки степени различия масштабированных данных и неприменимы в случае сравнения степени их различия.

3) Некоторые формулы неприменимы, ввиду использования при математическом масштабировании и нормировании делителя, не являющегося общим для сравниваемых и (или) совместно обрабатываемых числовых значений.

Учет обнаруженных недостатков позволит избежать ошибочных решений, обусловленных использованием неприменимых формул при решении задач в теории и практике экономики, организационного управления, медицины и многих других областей.

Библиографический список

1. Черноруцкий, И. Г. Методы принятия решений / И. Г. Черноруцкий. — Санкт-Петербург : БХВ-Петербург, 2005. — 416 с.

2. Кириченко, А. А. Нейропакеты — современный интеллектуальный инструмент исследователя [Электронный ресурс] / А. А. Кириченко. — Режим доступа: https://www.hse.ru/pubs/share/direct/document/91940629.pdf (дата обращения: 09.02.18).

3. Маршаков, Д. В. Нейросетевая идентификация динамики манипулятора [Электронный ресурс] / Д. В. Маршаков, О. Л. Цветкова, А. Р. Айдинян // Инженерный вестник Дона. — 2011. — № 3. — Режим доступа: http://www.ivdon.ru/magazine/archive/n3y2011/504 (дата обращения: 09.02.18).

4. Raschka, S. Python Machine Learning / S. Raschka. — Birmingham : Packt Publishing Ltd, 2015. — 454 p.

5. Россошанский, П. В. Анализ методов нормирования показателей качества сложных технических систем / П. В. Россошанский, С. А. Грайворонский // Надежность и качество : тр. междунар. симпозиума. — Пенза : Изд-во ПГУ. — 2010. — Т. 2, вып. 3. — С. 418–420.

6. Gnedenko, B. V. An elementary introduction to the theory of probability / B. V. Gnedenko, A. Ya. Khinchin. — Berlin : [s. n.], 2015. — 96 p.

7. Вентцель, Е. С. Теория вероятностей / Е. С. Вентцель. — Москва : КноРус, 2016. — 664 с.

8. Фатхи, В. А. Модели и алгоритмы диагностирования нейросетевых систем на основе модифицированных сетей Петри / В. А. Фатхи, Д. В. Маршаков, А. И. Зотов. — Ростов-на-Дону : Изд. центр ДГТУ, 2014. — 158 с.

9. Сидоров, А. А. Методические подходы к оценке эффективности деятельности органов государственной власти и местного самоуправления / А. А. Сидоров // Доклады ТУСУРа. — 2014. — № 1 (31). — С. 209–216.

10. Кузнецов, С. А. Большой толковый словарь русского языка / С. А. Кузнецов. — Санкт-Петербург : Норинт, 2000. — 1536 с.

11. Ожегов, С. И. Словарь русского языка / С. И. Ожегов. — Москва : Мир и образование, 2016. — 1376 с.

12. Долгов, А. И. Байесовские соотношения и их модификации / А. И. Долгов. — Ростов-на-Дону : Изд. центр ДГТУ, 2015. — 112 с.

13. Долгов, А. И. О корректности модификаций формулы Байеса / А. И. Долгов // Вестник Дон. гос. техн. ун-та. — 2014. — Т. 14, № 3 (78). — С. 13–20.

14. Naylor, C. M. Build Your Own Expert System / C. M. Naylor. — Chichester : John Wiley & Sons, 1987. — 289 p.

15. Бэстенс, Д.-Э. Нейронные сети и финансовые рынки: принятие решений в торговых операциях / Д. Э. Бэстенс, В.-М. Берг Ван Ден, Д. Вуд. — Москва : ТПВ, 1997. — 236 с.

References

1. Chernorutskiy, I.G. Metody prinyatiya resheniy. [Decision technique.] St.Petersburg: BKhV-Peterburg, 2005, 416 p. (in Russian).

2. Kirichenko, A.A. Neyropakety — sovremennyy intellektual'nyy instrument issledovatelya. [Neuro packages – a modern intellectual tool of the researcher.] Available at: https://www.hse.ru/pubs/share/direct/document/91940629.pdf (accessed: 09.02.18) (in Russian).

3. Marshakov, D.V., Tsvetkova, O.L., Aidinyan, A.R. Neyrosetevaya identifikatsiya dinamiki manipulyatora. [Neural network identification of dynamics of the manipulator.] Engineering Journal of Don, 2011, no. 3. Available at: http://www.ivdon.ru/magazine/archive/n3y2011/504 (accessed: 09.02.18) (in Russian).
4. Raschka, S. Python Machine Learning. Birmingham: Packt Publishing Ltd, 2015, 454 p.
5. Rossoshanskiy, P.V., Gaivoronskiy, S.A. Analiz metodov normirovaniya pokazateley kachestva slozhnykh tekhnicheskikh system. [Normalization principles of quality indices of complex engineering systems.] Nadezhnost' i kachestvo: tr. mezhdunar. Simpoziuma. [Reliability and quality: Proc. Int. Symposium.] Penza: PSU Publ. House, 2010, vol. 2, iss. 3, pp. 418–420 (in Russian).
6. Gnedenko, B.V., Khinchin, A.Ya. An elementary introduction to the theory of probability. Berlin: [s. n.], 2015, 96 p.
7. Ventsel, E.S. Teoriya veroyatnostey. [Probability Theory.] Moscow: KnoRus, 2016, 664 p. (in Russian).
8. Fatkhi, V.A., Marshakov, D.V., Zotov, A.I. Modeli i algoritmy diagnostirovaniya neyrosetevykh sistem na osnove modifitsirovannykh setey Petri. [Models and algorithms for diagnosing neural network systems based on modified Petri nets.] Rostov-on-Don: DSTU Publ. Centre, 2014, 158 p. (in Russian).
9. Sidorov, A.A. Metodicheskie podkhody k otsenke effektivnosti deyatel'nosti organov gosudarstvennoy vlasti i mestnogo samoupravleniya. [Methods for evaluation the effectiveness of public administration activities.] Proceedings of TUSUR University, 2014, no. 1 (31), pp. 209–216 (in Russian).
10. Kuznetsov, S.A. Bol'shoy tolkovyy slovar' russkogo yazyka. [The Great Explanatory Dictionary of the Russian language.] St.Petersburg: Norint, 2000, 1536 c. (in Russian).
11. Ozhegov, S.I. Slovar' russkogo yazyka. [Dictionary of the Russian language.] Moscow: Mir i obrazovanie, 2016, 1376 p. (in Russian).
12. Dolgov, A.I. Bayesovskie sootnosheniya i ikh modifikatsii. [Bayesian relations and their modifications.] Rostov-on-Don: DSTU Publ. Centre, 2015, 112 p. (in Russian).
13. Dolgov, A.I. O korrektnosti modifikatsiy formuly Bayesa. [On correctness of Bayes formula modifications.] Vestnik of DSTU, 2014, vol. 14, no. 3 (78), pp. 13–20 (in Russian).
14. Naylor, C. M. Build Your Own Expert System. Chichester: John Wiley & Sons, 1987, 289 p.
15. Bastans, D.-E., Berg Van Den, V.-M., Wood, D. Neyronnye seti i finansovye rynki: prinyatie resheniy v torgovykh operatsiyakh. [Neural networks and financial markets: decision-makings in commercial operations.] Moscow: TPV, 1997, 236 p. (in Russian)

Поступила в редакцию 19.11.2017
Сдана в редакцию 19.11.2017
Запланирована в номер 20.01.2018

Received 19.11.2017
Submitted 19.11.2017
Scheduled in the issue 20.01.2018

Об авторах:

Долгов Александр Иванович,
ведущий научный сотрудник АО «Всероссийский научно-исследовательский институт «Градиент» (РФ, 344010, г. Ростов-на-Дону, пр. Соколова, 96), доктор технических наук, профессор,
ORCID: http://orcid.org/0000-0001-8409-3467
dolgov-ai@yandex.ru

Маршаков Даниил Витальевич,
доцент кафедры «Вычислительные системы и информационная безопасность» Донского государственного технического университета (РФ, 344000, г. Ростов-на-Дону, пл. Гагарина, 1), кандидат технических наук,
ORCID: http://orcid.org/0000-0001-5795-8146
daniil_marshakov@mail.ru

Authors:

Dolgov, Alexander I.,
senior research scholar, All-Russian Scientific Research Institute "Gradient" JSC (RF, 344010, Rostov-on-Don, Sokolov Avenue, 96), Dr.Sci. (Eng.), professor,
ORCID: http://orcid.org/0000-0001-8409-3467
dolgov-ai@yandex.ru

Marshakov, Daniil V.,
associate professor of the Computer Systems and Information Security Department, Don State Technical University (RF, 334000, Rostov-on-Don, Gagarin sq., 1), Cand.Sci. (Eng.),
ORCID: http://orcid.org/0000-0001-5795-8146
daniil_marshakov@mail.ru